2009

Automatic Summarization of Customer Reviews: An Integrated Approach

Wingyan Chung
wchung@scu.edu

Automatic Summarization of Customer Reviews: An Integrated Approach

Wingyan Chung
Santa Clara University
wchung@scu.edu

ABSTRACT

The proliferation of interactivity between Web content producers and consumers underscores the development of the Internet in recent years. In particular, customer reviews posted on the Web have grown significantly. Because customers represent the primary stakeholder group of a company, understanding customers' concerns expressed in these reviews could help marketers and business analysts to identify market trends and to provide better products and services. However, the large volume of textual reviews written in informal language makes it difficult to understand customers' concerns. This paper describes an integrated approach to summarizing customer reviews. The approach consists of the steps of sentence extraction, aspect identification, sentiment classification, and review summarization. We report preliminary results of using our approach to summarize product reviews extracted from Amazon.com. Our work augments existing work by considering non-standard input and by incorporating linguistic resources and clustering in automatic summarization.

Keywords

Text summarization, online review, business intelligence, clustering, aspect analysis, e-commerce, sentiment analysis.

INTRODUCTION

As the Internet supports higher interactivity among users, customer reviews posted on the Web are growing significantly in recent years. The current Internet (Web 2.0) has evolved from its previous generation so that users are not only consumers of the Web content, but also are producers, oftentimes generating reviews on products and services. These reviews are read by people all over the world, producing direct influence on their sales and image. Understanding customers' sentiments and concerns expressed in these reviews is therefore important for companies to improve its products and services. However, the large volume of textual reviews posted on Web sites such as Amazon.com makes it difficult for marketers and business analysts to understand customers' concerns. Because customers are a major group of stakeholders of a company (Freeman, 1984), proper management and analysis of their concerns expressed in customer reviews is important. Decision makers can gain useful insights from the business intelligence revealed from these reviews (Weier and Smith, 2007). Unfortunately, existing work does not adequately address these issues. The use of informal language in these reviews further aggravates the difficulty in analyzing them.

In this paper, we review related works and describe an integrated approach to automatic summarization of a large number of customer reviews. The approach consists of the steps of sentence extraction, aspect identification, sentiment classification, and review summarization. We report preliminary results of using our approach to summarize product reviews extracted from Amazon.com. This research will contribute to the fields of text mining, electronic commerce, and human computer interaction by enriching the methods and techniques for summarizing customer reviews. It also will provide insights for customer relationship management and stakeholder analysis.

LITERATURE REVIEW

Analyzing customer reviews involves understanding many aspects, including customers' relationships with a company, their choices of words and expressions in reviews, and techniques for presenting review summaries. Because customers are a major stakeholder group of a company, literature on stakeholder management can inform the study of customer reviews. The field of text mining provides techniques for handling a large number of textual reviews. We review literature of these areas as follows.

Stakeholder Management

Previous work in stakeholder management studies corporate responsibilities on various stakeholders. Drawing from results of 78 field studies of corporate social performance (CSP) in major Canadian companies between 1983 and 1993, Clarkson (1995) proposed a framework for analyzing and evaluating CSP. Having mainly been conducted in the 1980s, the research does not consider more complex relationships in the e-commerce environment that has now evolved to be customer-centric. The stakeholder typology proposed by Mitchell, Agle and Wood (1997) identifies stakeholders by combinations of stakeholder attributes (power, legitimacy, urgency) but does not provide an automated system for summarizing customer reviews. Scholl (2001) reviewed the potential applcability of stakeholder theory to e-government but does not to how summarization of customer reviews can be facilitated on the Web. Chung, Chen and Reid (2009) developed a prototype called Business Stakeholder Aalyzer, which focuses on automatic classification of stakehholders. However, their system is not suitable for automatic summarization of product reviews. In summary, prior work on stakeholder management does not provide a systematic method for automatically summarizing stakeholders' (especially customers') reviews.

Text Mining for Handling Customer Reviews

Text mining has been applied to handling customer reviews. These applications try to analyze customer reviews to help recommend products, to calculate the utility of the reviews, to identify key product features, to detect spam, and to summarize review content. The goals often include better capability to market products and better understanding of customers' concerns. Zhang (2008) proposes new methods to weigh the customer ratings of different reviews. They use lexical similarity, shallow syntactic features, and lexical subjectivity clues to distinguish useful from useless reviews (usefulness is computed by how other people voted on a review). While their work addresses the reviews' usefulness that was not widely studied before, they do not provide any methods for summarizing the reviews.

Archak, Ghose and Ipeirotis (2007) adapt an econometric method called hedonic regression to estimate the weight that customers place on different product features and to calculate the extent that customers' implicit evaluation of these features affects the product's revenue. They show that textual portion of the reviews can improve product sales prediction compared to a baseline technique that simply relies on numeric data. But their work falls short of providing a useful summary of voluminous customer reviews.

To address ambiguity in review text, Ding and Liu (2007) use linguistic rules to determine the semantic orientations of words in customer reviews. While their rules can be used to aggregate opinion words about product features, they rely mainly on observing conjunction words, such as "but" and "however" and on looking for other clues showing semantic orientations in review sentences. Such observations may be particularly limited in short reviews commonly found on Web sites. Nevertheless, the work highlights the importance of incorporating linguistic resources in analyzing customer reviews.

Besides the ambiguous and interdependent nature of review text, spam also hinders understanding of customer reviews. Jindal and Liu (2007) categorize review spam into three types: false opinion (overly positive or negative comments), brand reviews (based only on brand but not product), and non-reviews (advertisements without comment). Their system identified duplicate reviews from different user IDs on the same product, duplicate reviews from the same user ID on different products, and duplicate reviews from different user IDs on different products. However, the system does not help readers to summarize large number of reviews.

Aciar, Zhang, Simoff and Debenham (2007) uses prioritized consumer product reviews to make recommendations. Using text-mining techniques, it maps automatically each sentence of a review into an ontology, which is a manually-created ontology for a product. While their system produces recommendations from textual reviews, its ontology can be used in analyzing only one product and is not adaptable to other products or domains.

Ling, Mei, Zhai and Schatz (2008) propose a system which allows the users to flexibly describe the information they are seeking for an arbitrary topic and generate a multi-faceted overview. They propose a two-stage framework working with probabilistic models. Zhuang, Jing and Zhu (2006) consider differet actors in summarizing movie reviews and used WordNet and statistical analysis in creating summaries. While these works attempt to summarize customer reviews, the performance could be much improved by considering a number of useful resources such as linguistic resources and product feature descriptions.

In summary, previous works on analyzing customer reviews lack consideration on helping marketers and business analysts to understanding large number of textual reviews. While some works exist on automatic summarization of customer reviews, they lack a comprehensive consideration of product features, textual semantic orientation, and the ambiguity and non-standard writing of online reviews. For example, spelling mistakes, use of acronyms, and short forms are found commonly in these reviews.

PROPOSED APPROACH AND PRELIMINARY RESULTS

We have developed an integrated approach to automatically summarize a large number of customer reviews posted on Web sites. The approach applies domain-specific heuristics, linguistic rules, and clustering algorithm to sentence extraction, sentiment classification, and review summarization.

Steps in the Approach

The steps of our approach are as follows.

Sentence Extraction

Like previous work, our work considers each sentence as an entity for expressing ideas and opinions. But unlike previous work, our approach takes non-standard textual input (online reviews written in informal language by non-professionals) and incorporates heuristics to precisely identify sentences from noisy review text. Sentence boundary determination for online consumer reviews is complicated due to the fact that these reviews are written in varied styles and they do not necessarily follow the conventional grammar. One important task is to recognize automatically when a period "." does not signify the end of a sentence. To recognize an End Of Sentence (EOS), we modified the sentence extraction algorithm proposed in (Weiss, Indurkhya, Zhang and Damerau, 2005) by incorporating the following new guidelines:

- All '?' and '!' are EOS
- If '"' or '' appears before period, it is EOS
- If ')',')' or ']' appears before period, it is EOS
- If period is followed by other periods, it is not EOS
- If token before period is a digit and token following period is a digit, it is not EOS
- If token before period is 'Mr', 'Ms', 'Miss', 'Mrs', 'Dr', it is not EOS
- If token following period is 'com', 'org', it is not EOS
- If token before period is 'O' and token after period is 'S', it is not EOS
- If token before period is 'U' and token after period is 'S', it is not EOS
- If token before period is 'i' and token after period is 'e', it is not EOS
- If token before period is 'e' and token after period is 'g', it is not EOS

Aspect Analysis

Our approach identifies various aspects expressed in a review by considering the keyword distribution in all the reviews related to a product. Each aspect is represented by a set of unique key words. All stop words are removed and word stemming is applied in the calculation. For each sentence extracted from the previous step, we calculate the relevance score between the sentence and an aspect and categorize a sentence into an aspect with the highest aspect score.

The calculation of aspect score is based on an external aspect keyword file that contains a list of aspects that we identified manually from a random sample of customer reviews. The aspects of each product are listed in Table 2. Each aspect is characterized by a list of keywords. Each word from the extracted sentences is compared to the aspect keyword list and a score is calculated as follows:

- A counter is maintained for each of the aspects, every time a matching keyword is found in the sentence, the counter for that aspect is incremented by 1.
- For each sentence, the aspect with the highest counter value is identified.
- In order to normalize the aspect score, the counter value is divided by the total number of keywords for that particular aspect in the aspect keyword file.

Sentence Aspect Score(x) = (Number of matches(x))/L

where

Number of matches(x) is the total number of words in the sentence that match the keywords defined for the aspect x,

L is total number of keywords for aspect x

Sentiment Analysis

We use a list of 8,221 opinion words (Wiebe, Wilson and Cardie, 2005) to calculate the sentiment scores of the extracted sentences, with a view to identify strong opinions from users' reviews. The word list provides information on the strength (strongly or weakly subjective) and polarity (positive or negative) of the sentiment expressed by the words. Each word from a sentence is compared to the sentiment word list and a score is calculated as follows:

- If a word is positive and strongly subjective, sentiment score for the sentence is incremented by 1.0
- If a word is negative and strongly subjective, sentiment score for the sentence is decremented by 1.0

- If a word is positive and weakly subjective, sentiment score for the sentence is incremented by 0.5
- If a word is negative and weakly subjective, sentiment score for the sentence is decremented by 0.5

Sentence Sentiment Score = $\left(\sum_{i=1}^{L} Sentiment\_Score(i) \right) / L$

where

Sentiment_Score(i) is the sentiment score of word i calculated based on above mentioned rules,

L is total number of non-stop words in sentence

Each sentence's sentiment score is then normalized so that sentence length does not adversely affect the calculation. These scores are used to extract opinions that reflect strong sentiment and may require attention of marketers and business analysts.

Review Clustering

Our approach uses Multidimensional Dimensional Scaling algorithms to provide map-like views for the summaries (Torgerson, 1952). Two levels of summary are provided: A sentence-level summary displays each sentence as a point on the map while a review-level summary displays each review as a point. As the algorithm also clusters these objects by their similarity, users are able to visually compare the outputs of the algorithms and to save time understanding the review content.

To our knowledge, no existing work supports automatic summarization of customer reviews using an integrated approach such as ours. Our work differs from existing works in that we consider non-standard input (text segments written by non-professionals and in informal language) and that we incorporate linguistic resources and clustering in the summarization process, thus producing richer summaries that enhance human understanding.

Preliminary Results

To study the feasibility of our proposed approach, we developed a research test bed consisting of product reviews posted on Amazon.com Web site. Amazon Java API was used for extracting the reviews. Products on the Amazon Web site have a large number of reviews for wide range of products. Each review includes a title, a text review, date, time, author name and location, ratings, and other miscellaneous information. We have tested our approach using the customer reviews of five products: Acer Aspire One 8.9-inch Mini Laptop, Apple MacBook MB402LL/B, HP 2133-KR922UT 8.9-Inch Mini-Note PC, PlayStation 3 Dualshock 3 Wireless Controller, and Radio Flyer Deluxe Steer tricycle. Table 1 shows the basic information about the reviews and the results of sentence extraction.

Sentence Extraction

Sentences were extracted from reviews for each of the five products listed above. The sentences were sampled on a systematically random basis (1 per every 10 sentences) to determine the sentence extraction accuracy.

Analysis Metrics	Product				
	Acer Mini laptop	Apple MacBook	HP Mini-Note PC	Playstation Wireless Controller	Radio Flyer Tricycle
Total number of reviews	521	213	113	252	132
Total number of sentences extracted	6135	2340	1283	1194	971
Number of sentences sampled	613	234	129	119	98
Sentence extraction accuracy	94.78%	94.87%	97.67%	96.63%	95.9%

Table 1. Results of Sentence Extraction

Aspect and Sentiment Analyses

Four to six aspects were identified based on manually reading a random sample of customer reviews for each product. Then keywords of these aspects were extracted for calculating aspect scores. These aspects are shown in Table 2.

Product				
Acer Mini laptop	**Apple MacBook**	**HP Mini-Note PC**	**Playstation Wireless Controller**	**Radio Flyer Tricycle**
Body	Specification	Body	Specification	Body
Performance	Body	Performance	Appearance	Durability/safety
Purchase	Performance	Purchase	Purchase	Purchase
Specification	Manufacturer	Case		Manufacturer
	Purchase			

Table 2. Aspects of Each Product

Table 3 shows for each product the sentences extracted that have the highest and lowest sentiment scores and also the aspects they belong to.

Product	Sentiment	Sentence	Sentiment Score	Aspect
Acer Mini laptop	Sentence with the Most Positive Sentiment	"Awesome."	1.00	Unidentified
	Sentence with the Most Negative Sentiment	"The mouse is annoying as heck."	-0.571	Unidentified
Apple MacBook	Sentence with the Most Positive Sentiment	"Hope."	1.00	Purchase
	Sentence with the Most Negative Sentiment	"Big mistake."	-0.667	Specification
HP Mini-Note PC	Sentence with the Most Positive Sentiment	"Wow this thing is amazing."	0.8	Unidentified
	Sentence with the Most Negative Sentiment	"Don't bother."	-0.667	Unidentified
Playstation Wireless Controller	Sentence with the Most Positive Sentiment	"it's perfect."	1.00	Unidentified
	Sentence with the Most Negative Sentiment	"Seriously."	-0.1	Unidentified
Radio Flyer Tricycle	Sentence with the Most Positive Sentiment	"That gets awesome."	0.5	Unidentified
	Sentence with the Most Negative Sentiment	"I regret this purchase."	-0.75	Purchase

Table 3. Results of Sentence Extraction

Discussion

Several observations are found in these preliminary findings. First, the sentence extraction accuracies across all tested product are very high (> 94.7%), showing promise of our approach in handling noisy product reviews. Second, the aspect of many extracted sentences was unidentified because many keywords in the sentences do not appear in our lists of keywords. Third, the sentences extracted reflect the positive and negative sentiment in the reviews. There is a need to conduct further evaluation to study the effectiveness of the approach. Currently we are in the process of applying and evaluating the MDS algorithm to analyzing the reviews. A preliminary result of the MDS clustering is shown in Figure 1, which shows some of the reviews of the Sony Playstation and their relationships. Each review is represented by a node and each line represents a relationship between two nodes. The MDS algorithm clusters reviews based on their similarity such that similar reviews are placed closer to each others.

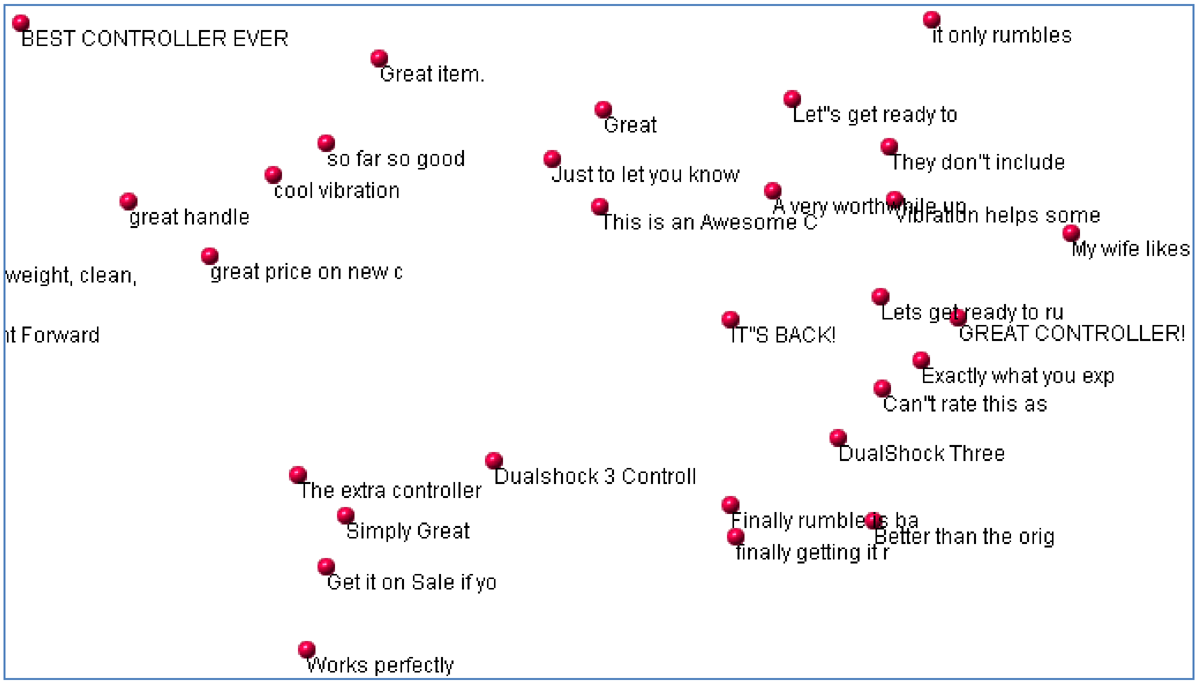

Figure 1. Summarizing reviews for Sony Playstation using the MDS algorithm

CONCLUSION

As the Internet supports higher interactivity among users, customer reviews posted on the Web are growing significantly in recent years. However, the large volume of customer reviews posted on Web sites such as Amazon.com makes it difficult for marketers and business analysts to understand customers' concerns. In this paper, we describe an approach to automatic summarization of customer reviews and present preliminary results of our study of reviews on products listed on Amazon.com. The findings demonstrate a generally high accuracy in extract sentences from noisy customer reviews. Positive and negative sentiments were identified correctly. We are now in the process of conducting a user evaluation to confirm the effectiveness of our results. We believe that our work will contribute to better techniques and understanding of summarizing customer reviews and will benefit online marketers, business intelligence practitioners, and researchers in the fields of text mining and e-commerce.

ACKNOWLEDGMENTS

This research was partly supported by the 2009 Presidential Research Grant provided by Santa Clara University. We thank the track chairs and anonymous reviewers for their work.

REFERENCES

1. Aciar, S., Zhang, D., Simoff, S. and Debenham, J. (2007) "Informed recommender: basing recommendations on consumer product reviews," *IEEE Intelligent Systems*, 22, 3, 39-47.
2. Archak, N., Ghose, A. and Ipeirotis, P.G. (2007) "Show me the money!: deriving the pricing power of product features by mining consumer reviews," *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM New York, NY, USA, San Jose, California, USA, 56-65.
3. Chung, W., Chen, H. and Reid, E.O.F. (2009) "Business stakeholder analyzer: An experiment of classifying stakeholders on the Web," *Journal of the American Society for Information and Science and Technology*, 60, 1, 59-74.
4. Clarkson, M.B.E. (1995) "A stakeholder framework for analyzing and evaluating corporate social performance," *Academy of Management Review*, 20, 1, 92-117.

5. Ding, X. and Liu, B. (2007) "The utility of linguistic rules in opinion mining," *Proceedings of the 30th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, Amsterdam, The Netherlands, 811-812.
6. Freeman, E. (1984) *Strategic Management: A Stakeholder Approach* Pitman, Marshfield, MA.
7. Jindal, N. and Liu, B. (2007) "Analyzing and detecting review spam," *Proceedings of the 7th IEEE International Conference on Data Mining*, IEEE Press, Omaha, NE, 547-552.
8. Ling, X., Mei, Q., Zhai, C. and Schatz, B. (2008) "Mining multi-faceted overviews of arbitrary topics in a text collection," *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, Las Vegas, Nevada, USA, 497-505.
9. Mitchell, R.K., Agle, B.R. and Wood, D.J. (1997) "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts," *Academy of Management Review*, 22, 4, 853-886.
10. Scholl, H.J. (2001) "Applying stakeholder theory to e-government: benefits and limits," *Proceedings of the First IFIP conference on e-Commerce and e-Government*, Zurich, Switzerland.
11. Torgerson, W.S. (1952) "Multidimensional scaling: I. Theory and Method," *Psychometrika*, 17, 4, 401-419.
12. Weier, M.H. and Smith, L. (2007) "Many Companies Plan To Increase BI Spending," *InformationWeek*, 2007/03/17.
13. Weiss, S., Indurkhya, N., Zhang, T. and Damerau, F. (2005) *Text Mining: Predictive Methods for Analyzing Unstructured Information* Springer, New York, NY.
14. Wiebe, J., Wilson, T. and Cardie, C. (2005) "Annotating expressions of opinions and emotions in language," *Language Resources and Evalution (formerly Computers and the Humanities)*, 1, 2, 1-54.
15. Zhang, Z. (2008) "Weighing stars: Aggregating online product reviews for intelligent e-commerce applications," *IEEE Intelligent Systems*, 23, 5, 42-49.
16. Zhuang, L., Jing, F. and Zhu, X.-Y. (2006) "Movie review mining and summarization," *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, ACM Press, Arlington, VA, 43-50.